PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06T 15/10 — A1

(11) International Publication Number: WO 99/06956

(43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/IB98/00983

(22) International Filing Date: 25 June 1998 (25.06.98)

(30) Priority Data:
97401822.8 29 July 1997 (29.07.97) EP
(34) Countries for which the regional or international application was filed: FR et al.

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL–5621 BA Eindhoven (NL).

(71) Applicant *(for SE only)*: PHILIPS AB [SE/SE]; Kottbygatan 7, Kista, S–164 85 Stockholm (SE).

(72) Inventor: DUFOUR, Cécile; Prof. Holstlaan 6, NL–5656 AA Eindhoven (NL).

(74) Agent: LANDOUSY, Christian; Internationaal Octrooibureau B.V., P.O. Box 220, NL–5600 AE Eindhoven (NL).

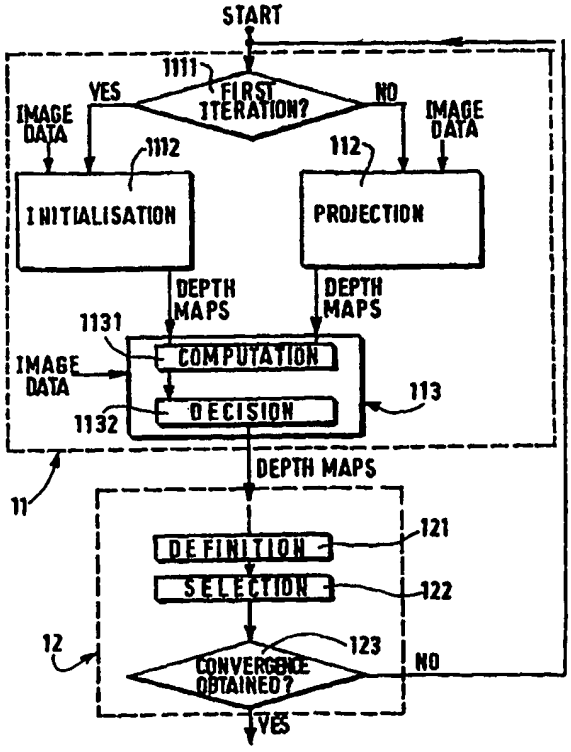
(81) Designated States: JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD OF RECONSTRUCTION OF TRIDIMENSIONAL SCENES AND CORRESPONDING RECONSTRUCTION DEVICE AND DECODING SYSTEM

(57) Abstract

The invention relates to a new method of reconstruction of tridimensional scenes. While conventional methods are often limited to the 3D reconstruction of the bounding volume of the concerned objects, the proposed method of recovery of a 3D geometric model from 2D views taken by one single camera, giving an information even about the parts which are hidden in each view, is implemented according to a first depth labeling step, implemented in a sub–system (11) and including initialisation and projection sub–steps followed by a refinement process, and to a second reconstruction step, implemented in a sub–system (12). By means of a close cooperation of the 3D depth maps thus obtained for two views of a scene, a 3D model is identified and extracted. Application = additional fonctionality in multimedia services.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Method of reconstruction of tridimensional scenes and corresponding reconstruction device and decoding system.

The present invention relates to a method of reconstruction of a tridimensional scene from a bidimensional video sequence corresponding to N successive images of a real scene, and to a corresponding reconstruction device and a decoding system.

In light of recent advances in technology (and in the framework of all what is related to the future MPEG-4 standard intended to provide means for encoding graphic and video material as objects having given relations in space and time) all what relates to stereo images and virtual environments is becoming an important tool, for instance in engineering, design or manufacturing. Stereo images, usually generated by recording two slightly different view angles of the same scene, are perceived in three dimensions (3D) if said images are considered by pairs and if each image of a stereo pair is viewed by its respective eye. Moreover, in such stereo and virtual reality contexts, a free walkthrough into the created environments is required and possible. This creation of virtual environments is performed by means of picture synthesis tools, typically according to the following steps :

(a) a recovery step of a 3D geometric model of the concerned scene (for instance, by using a facet representation) ;

(b) a rendering step, provided for computing views according to specific points of view and taking into account all the known elements (for instance, lights, reflectance properties of the facets, correspondence between elements of the real views,...)

The reconstruction of a 3D geometric model of a scene however requires to perform an image matching among all available views. In the document "Multiframe image point matching and 3D surface reconstruction", R.Y. Tsai, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.PAMI-5, n°2, March 1983, pp.159-174, such a correspondence problem is solved by computing a correlation function that takes into account (inside a defined search window, along an axis corresponding to the sampling grid of the input pictures) the information of all the other views in one single pass, providing in this way a rather robust method against noise and periodical structures. The minimum of this function provides an estimate of the depth of the pixel in the center of the search window. Unfortunately, this depth estimate has a non-linear dependence (in 1/x in the most simple case) to the sampling grid. Moreover, the depth map estimation for a surface obtained from one picture cannot be easily

compared with the depth map estimation of the same surface obtained from another picture, because they do not share the same reference grid (they are only referenced to their respective picture sampling grid).

A first object of the invention is to propose a scene reconstruction method which no longer shows these drawbacks.

To this end the invention relates to a method of reconstruction such as defined in the preamble of the description and which is moreover characterized in that it comprises in series, for each image, segmented into triangular regions, of the sequence :

(A) a first depth labeling step, in which, each view being considered as the projection of a continuous 3D sheet, a multi-view matching is performed independently on each view in order to get a disparity map corresponding to the depth map of said 3D sheet ;

(B) a second 3D model extraction step, in which an octree subdivision of the 3D space is performed and the voxels (volume elements) lying in the intersection of all 3D depth sheets are kept. An octree is a tree-structured representation used to describe a set of binary valued volumetric data enclosed by a bounding cube and constructed by recursively subdividing each cube into eight subcubes, starting at the root node which is a single large cube : octrees are an efficient representation for many volumetric objects since there is a large degree of coherence between adjacent voxels in a typical object.

With such an approach, a correlation function along an axis corresponding to sampled values of depth in the 3D world coordinates system (constituting a depth sampling grid provided at will by the user) is computed taking all views into account, and the minimum of this function is directly related to an accurate value of depth in said coordinates system (this is a great advantage when multiple depth estimations are obtained from different viewpoints). The depth sampling grid is provided by the user at will and is advantageously chosen regularly spaced, taking however into account some preliminary knowledge about the surface to be reconstructed (for instance if said surface is known to lie within a predefined bound box, which is the case for indoor scenes).

The document USP 5598515 describes a system and method for reconstructing a tridimensional scene or elements of such a scene from a plurality of bidimensional images of said scene, but according to a complex procedure that is replaced, in the case of the invention, by a much more simple one submitted to successive refinements until convergence is obtained.

According to the invention, said depth labeling step preferably comprises in series an initialisation sub-step, provided for defining during a first iteration a preliminary 3D depth sheet for the concerned image, and a refinement sub-step, provided for defining, for each

vertex of each region, an error vector corresponding for each sampled depth to the summation of correlated costs between each of the (N-1) pairs of views (for a sequence of N images) on a window specifically defined for said vertex and storing the index that provides the minimum correlation cost, an additional operation being intended to replace after the first iteration the initialisation sub-step by a projection sub-step provided first for adjusting the position and field of view of the image acquisition device according to its parameters and the vertex map near to the image plane, and then for listing for each vertex the voxels that intersect the line passing through the vertex and the optical center of said acquisition device, in the viewing direction, and selecting the nearest voxel to the image plane. Concerning said 3D model extraction step, it preferably comprises in series a resolution definition sub-step, provided for defining the resolution of the voxel grid, and a voxel selection sub-step, provided for keeping for each view the voxels lying inside the non-empty spaces provided by each depth map and then only keeping voxels lying at the intersection of all non-empty spaces.

Another object of the invention is to propose a reconstruction device allowing to carry out this method.

To this end the invention relates to a device for reconstructing a tridimensional scene from a bidimensional video sequence corresponding to N successive images of a real scene, characterized in that :

(I) each of the N images of the sequence is segmented into triangular regions ;

(II) said device comprises, for processing each image of said sequence ;

(A) a depth labeling sub-system, comprising itself in series :

(1) an initialisation device, provided for defining during a first iteration an error vector corresponding for a set of sampled depths to the summation of correlation costs between each of the (N-1) pairs of views and the index providing the minimum correlation cost, the depth value of each vertex of the regions being computed by interpolation between the depths obtained for the neighboring regions ;

(2) a refinement device, provided for defining similarly for each vertex an error vector on a previously delimited window and, correspondingly, the index providing the minimum correlation cost ;

(B) a reconstruction sub-system provided for selecting the resolution of the voxel grid and keeping, for each view, the voxels lying inside the non-empty spaces provided by each depth map and, finally, only the voxels lying at the intersection of all non-empty spaces ;

(III) said depth labeling sub-system also comprises a projection device intended to replace during the following iterations the initialisation device and provided for adjusting the position and field of view of the image acquisition device, and the vertex map very near to the image plane, and, for each vertex, listing the voxels that intersect the line passing through the vertex and the optical center of said acquisition device in the viewing direction and selecting the nearest voxel to the image plane. The invention also relates to a video decoding system including such a reconstruction device.

The advantages of the invention will now be better understood by referring to the following description and the accompanying drawings, in which :

Fig. 1 shows the global scheme of a reconstruction device according to the invention ;

Fig. 2 illustrates the operations carried out in the initialisation device of the device of Fig. 1 ;

Fig. 3 illustrates the operations carried out in the refinement device of the device of Fig. 1 ;

Fig. 4 illustrates the operations carried out in the 3D reconstruction sub-system of the device of Fig. 1 ;

Fig. 5 illustrates the operations carried out in the projection device of the device of Fig. 1.

The device shown in Fig.1 is intended to allow, according to the invention, the reconstruction of scenes in three-dimensional form (3D), based on a sequence of N successive bidimensional images (2D) of said scenes. Said recovery is realized in two sub-systems 11 and 12, according to an implementation in two steps which are aimed to be iterated. The first step is a depth labeling one : each view is considered as the projection of a continuous 3D sheet, and a multi-view matching is performed independently on each view to get its disparity map, each disparity map then corresponding to the depth map of the 3D sheet (the disparity, the measurement of which provides a depth estimate, is the shift of a patch on the left (right) image relative to the right (left) image, and the output of any correspondence problem is a disparity map). The second step is a 3D model extraction one : an octree subdivision of the 3D space is performed and voxels lying in the intersection of all 3D depth sheets are kept.

The device of Fig.1 is therefore subdivided into two parts : the depth labeling sub-system 11, for carrying out the first depth labeling step, and the 3D reconstruction sub-system 12, for carrying out the second 3D model extraction step. The depth labeling sub-

system 11 itself comprises an initialisation device, a projection device 112, and a refinement device 113.

The initialisation device comprises, as illustrated in Fig.1, a test circuit 1111 followed by an initialisation circuit 1112. The test circuit 1111 is provided for switching either towards the circuit 1112 (YES) when the iteration is the first one, at the start of the procedure, or towards the device 112 (NO) when the initialisation has already been done.

If I is the image for which one wishes to recover a depth sheet and I1 to IN the pictures used for multi-view matching, it is supposed that, within the concerned field of view, I is segmented into triangular regions supposed to lie parallel to the image plane of I. For each region R(I), three operations are then successively carried out in sub-steps 1112a, 1112b and 1112c (illustrated in Fig.2), in order to obtain in the current field of view the depth of this region among a set S of predetermined depths D1, D2,..., Di,..., DM.

The sub-step 1112a (Fig.2, upper part) allows to compute for each region an error vector V(i) of defined length, said vector corresponding, for each sampled depth (C is the reference optical center), to the summation of correlation costs between each of the (N-1) pairs of views (image i, image j), which may be expressed by :

$$V(i) = \sum_{i=1}^{i=N} \mathrm{err}(i)[Ii, Ij]$$

Each coordinate i of V(i) corresponds to the sum of the errors encountered at depth Di in each view. The correlation measure err(i)[Ii, Ij] is a mean squared error between pixels of R(I) and pixels of the region R(Ij) in the image Ij assumed to lie at depth Di and obtained using the projection matrix relating the coordinates systems of the views I and Ij. The sub-step 1112b (Fig.2, middle part) allows to find for each region the index providing the minimum correlation cost, and the sub-step 1112c (Fig.2, lower part) to compute for each vertex of each region its depth value, by interpolation between the depths obtained for the neighboring regions (i.e. the depth of each vertex of the triangular regions will be the average of the depths of the regions sharing the vertex).

Thank to the initialisation, a preliminary 3D depth sheet is obtained for the image I. Each region R(I) has now an estimate of its 3D position and orientation, given by the 3D coordinates of its three vertices. However said orientation of the regions no longer complies with the initial assumption that they lie parallel to the image plane of the image I.

The initialisation device might then be used iteratively and run again while taking into account the new estimates of the orientations of each region in the image I. Another approach has finally been preferred : instead of searching for error vectors independently on

each region, error vectors are searched independently for the vertices in I (depth estimates are now searched for each vertex while leaving the depth estimates on neighboring vertices unchanged). This approach is carried out in the refinement device 113.

This device 113, which receives, as illustrated in Fig.1 (and in Fig.3 showing the sub-steps carried out in said device), the depth maps available at the output of the circuit 1112, first comprises a vector computation circuit 1131, in which, for each vertex, a window W on which correlation costs will be measured is defined (Fig.3, upper part). For each vertex, an error vector is then computed (Fig.3, middle part), that corresponds, for each sampled depth, to the summation of correlation costs between each of the (N-1) pairs of views (image i, image j) on the delimited window. In a decision circuit 1132, the index providing the minimum correlation cost for each vertex is then found (Fig.3, lower part). A refined 3D depth sheet is now available.

The depth maps available at the output of the device 113 are the output signals of the depth labeling sub-system 11 and are sent towards the 3D reconstruction sub-system 12, that comprises, as illustrated in Fig.1 (and in Fig.4 showing the sub-steps carried out in said device), a resolution definition device 121 followed in series by a voxel selection device 122 and a test circuit 123. In the device 121, the resolution of the voxel grid is chosen (Fig.4, upper part). In the device 122, for each view, the voxels lying inside the non-empty spaces provided by each depth map are kept (Fig.4, middle part), and only the voxels lying at the intersection of all non-empty spaces are finally kept (Fig. 4, lower part). A test of convergence is then done in the test circuit 123, some of the previous steps having to be iterated until said convergence is obtained.

As the initialisation has been done during the previously described first iteration, at the beginning of the second one the test circuit 1111 now switches towards the projection device 112. With respect to the first sub-steps 1112a, 1112b, 1112c carried out in the circuit 1112, the sub-steps 1121a, 1121b now provided in the device 112 and illustrated in Fig.5 allow : (a) to adjust (Fig.5, upper part) the position and the field of view of the camera according to the camera parameters, and the vertex map very near to the image plane, and : (b) to list (Fig.5, middle part) for each vertex the voxels that intersect the line passing through the vertex and the optical center of the camera, in the viewing direction, and to select the nearest voxel to the image plane. The output of said device 112, illustrated in Fig.5, lower part, is then sent (as the output of the device 112 in the case of the first iteration) towards the refinement device 113, that functions as already described.

CLAIMS:

1. A method of reconstruction of a tridimensional scene from a bidimensional video sequence corresponding to N successive images of a real scene, comprising in series, for each image, segmented into triangular regions, of the sequence :

(A) a first depth labeling step, in which, each view being considered as the projection of a continuous 3D sheet, a multi-view matching is performed independently on each view in order to get a disparity map corresponding to the depth map of said 3D sheet ;

(B) a second 3D model extraction step, in which an octree subdivision of the 3D space is performed and the voxels lying in the intersection of all 3D depth sheets are kept.

2. A method according to claim 1, wherein said depth labeling step comprises in series an initialisation sub-step, provided for defining during a first iteration a preliminary 3D depth sheet for the concerned image, and a refinement sub-step, provided for defining, for each vertex of each region, an error vector corresponding for each sampled depth to the summation of correlated costs between each of the (N-1) pairs of views on a window specifically defined for said vertex and storing the index that provides the minimum correlation cost, an additional operation being intended to replace after the first iteration the initialisation sub-step by a projection sub-step provided first for adjusting the position and field of view of the image acquisition device according to its parameters and the vertex map near to the image plane, and then for listing for each vertex the voxels that intersect the line passing through the vertex and the optical center of said acquisition device, in the viewing direction, and selecting the nearest voxel to the image plane.

3. A method according to claim 2, wherein said 3D model extraction step comprises in series a resolution definition sub-step, provided for defining the resolution of the voxel grid, and a voxel selection sub-step, provided for keeping for each view the voxels lying inside the non-empty spaces provided by each depth map and then only keeping voxels lying at the intersection of all non-empty spaces.

4. A device for reconstructing a tridimensional scene from a bidimensional video sequence corresponding to N successive images of a real scene, characterized in that :

(I) each of the N images of the sequence is segmented into triangular regions ;

(II) said device comprises, for processing each image of said sequence ;

(A) a depth labeling sub-system, comprising itself in series :

(1) an initialisation device, provided for defining during a first iteration an error vector corresponding for a set of sampled depths to the summation of correlation costs between each of the (N-1) pairs of views and the index providing the minimum correlation cost, the depth value of each vertex of the regions being computed by interpolation between the depths obtained for the neighboring regions ;

(2) a refinement device, provided for defining similarly for each vertex an error vector on a previously delimited window and, correspondingly, the index providing the minimum correlation cost ;

(B) a reconstruction sub-system provided for selecting the resolution of the voxel grid and keeping, for each view, the voxels lying inside the non-empty spaces provided by each depth map and, finally, only the voxels lying at the intersection of all non-empty spaces ;

(III) said depth labeling sub-system also comprises a projection device intended to replace during the following iterations the initialisation device and provided for adjusting the position and field of view of the image acquisition device, and the vertex map very near to the image plane, and, for each vertex, listing the voxels that intersect the line passing through the vertex and the optical center of said acquisition device in the viewing direction and selecting the nearest voxel to the image plane.

5. A video decoding system comprising a reconstruction device according to claim 4.

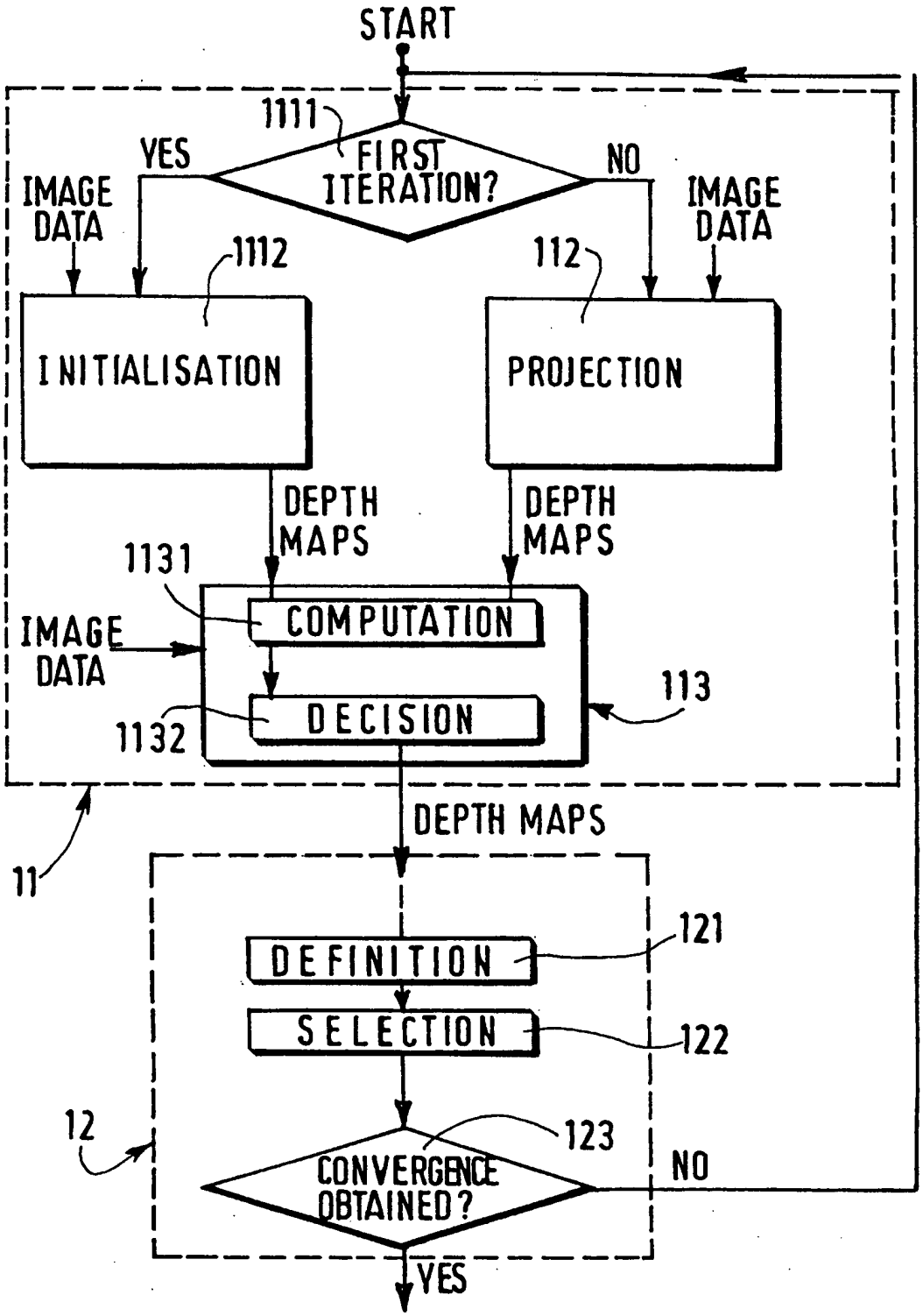

START
1111
FIRST ITERATION?
YES
NO
IMAGE DATA
IMAGE DATA
1112
112
INITIALISATION
PROJECTION
DEPTH MAPS
DEPTH MAPS
1131
COMPUTATION
IMAGE DATA
DECISION
1132
113
11
DEPTH MAPS
121
DEFINITION
SELECTION
122
12
123
CONVERGENCE OBTAINED?
NO
YES

FIG.1

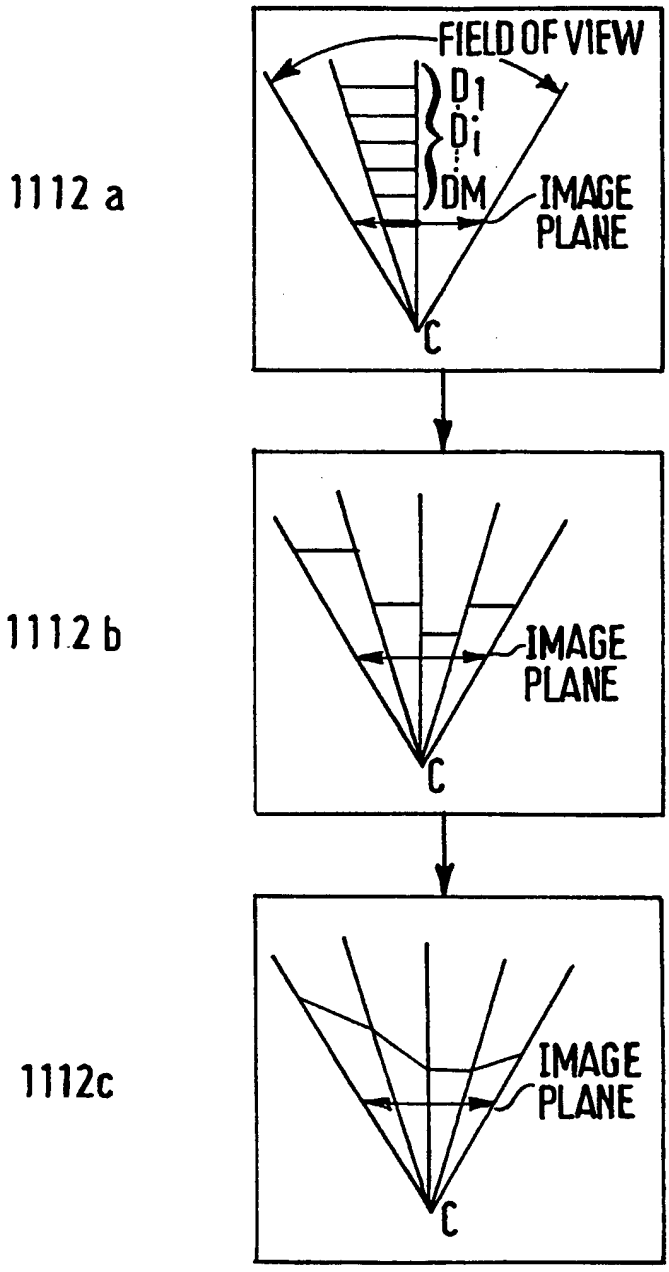

1112 a
FIELD OF VIEW
D1
Di
DM
IMAGE PLANE
C
1112 b
IMAGE PLANE
C
1112c
IMAGE PLANE
C

FIG.2

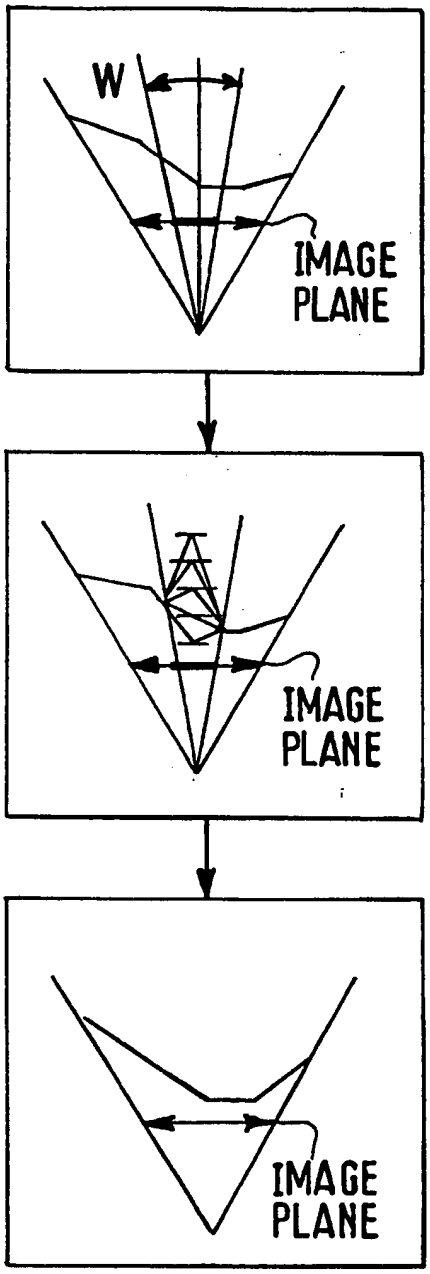


FIG.3

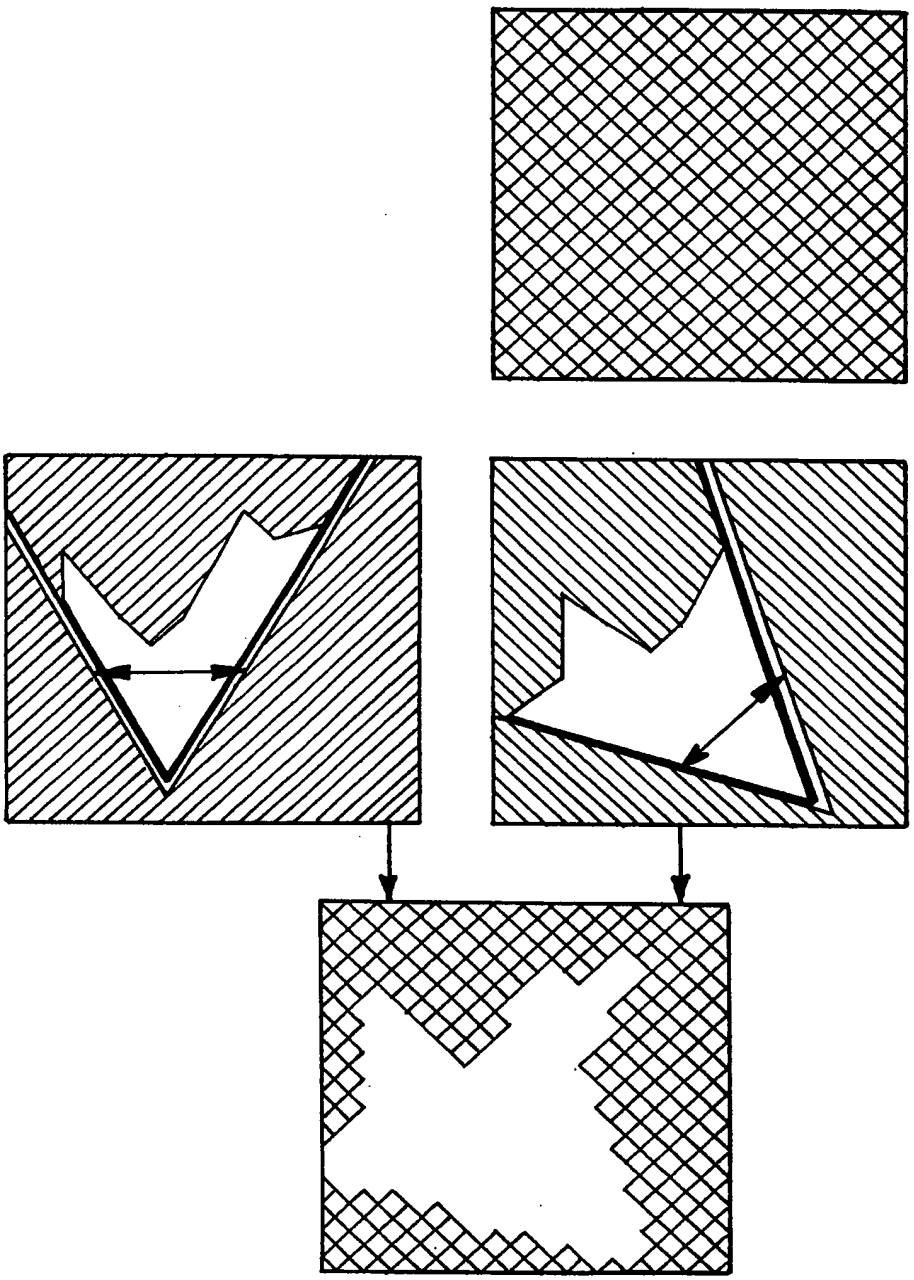


FIG.4

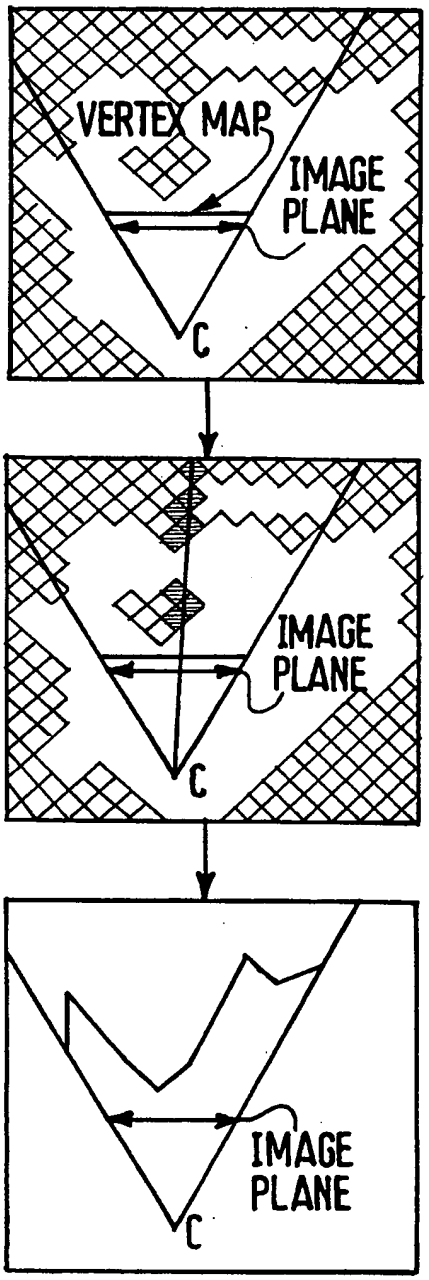


FIG.5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 98/00983

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: G06T 15/10
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Proceedings of the International Conference on Image Processing, Volume 3,September 16-19,1996, (Lausanne,Switzerland),F.Leymarie et al,"REALISE": Reconstruction of REALity from Image Sequences", page 651-page 654 --	1-5
A	Computer Graphics Proceedings 1996 (Siggraph), August 4-9, 1996,(New Orleans,USA),Paul E.Debevec et al,"Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach", page 11-page 20 --	1-5

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 December 1998	17 -12- 1998
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Malin Keijser Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB 98/00983

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Proceedings of the Fifth International Conference on Computer Vision, June 20-23,1995,(Cambridge, Massachusetts), Anders Heyden, " Reconstruction from Image Sequences by means of Relative Depths", page 1058 - page 1063 -- --------	1-5